

# When the Teacher Is the Test Proctor

*Ruth A. Childs & Linda Umezawa*

*Ontario Institute for Studies in Education (OISE)*

*University of Toronto*

This study investigates what grade-3 teachers say they would do if faced with common test administration dilemmas – and why. Grade-3 teachers with experience administering Ontario's provincially mandated assessment were recruited through professional association newsletters and Toronto-area newspapers. They responded to an on-line questionnaire ( $n = 98$ ) or an hour-long interview ( $n = 40$ ). Many teachers predicted that they would not follow the test administration instructions. Their rationales included (a) supporting the students to do their best work, (b) ensuring a positive testing experience for the students, and (c) maintaining their pedagogical routines and their relationships with the students. The findings highlight the ethical dilemmas teachers may experience when proctoring tests.

Key words: large-scale assessment, standardized testing, test administration, Ontario education

Les auteures analysent ce que des enseignants de 3<sup>e</sup> année feraient selon eux s'ils étaient confrontés avec les dilemmes courants reliés à l'administration de tests et les raisons pour lesquelles ils agiraient ainsi. Des enseignants de 3<sup>e</sup> année habitués à faire passer des tests provinciaux en Ontario ont été recrutés par le biais d'annonces dans des bulletins d'associations professionnelles et des journaux de la région de Toronto. Ils ont répondu à un questionnaire en ligne ( $n = 98$ ) ou participé à une entrevue d'une heure ( $n = 40$ ) ; près de 80 % des enseignants dans les deux groupes provenaient de la Région du Grand Toronto. De nombreux enseignants ont indiqué qu'ils ne suivraient pas les directives au sujet de l'administration des tests parce qu'ils veulent, entre autres : (a) aider les élèves à faire de leur mieux, (b) faire en sorte que les élèves aient une expérience positive des tests et (c) suivre leurs méthodes pédagogiques habituelles et conserver leurs liens avec les élèves. Les conclusions des auteures mettent en évidence les dilemmes éthiques auxquels les enseignants peuvent faire face lorsqu'ils surveillent l'administration de tests.

Mots clés : tests communs, tests normalisés, administration de tests, éducation en Ontario.

In many jurisdictions, elementary teachers not only prepare their students to take large-scale assessments, but also administer the assessments to their students. Much has been written about what teachers should and should not do to prepare their students to take large-scale assessments (e.g., Cohen & Hyman, 1991; Crocker, 2006; Haladyna, Nolen, & Hass, 1991; Mehrens, 1991; Mehrens & Kaminski, 1989; Mehrens, Popham, & Ryan, 1998; Popham, 1991). Less has been written about test administration, perhaps because what teachers should do in administering large-scale assessments is thought to be both unambiguous and uninteresting – to strictly follow the test administration instructions provided by the test developer. Not surprisingly, most research on test administration has focused on detecting deviations from the instructions (e.g., Cizek, 1999, 2003). The reasons teachers choose to follow – or not to follow – the instructions have received little attention.

This study investigates what teachers predict they would do if faced with common test administration dilemmas – and why. The study involved grade-3 teachers with experience administering Ontario's provincially mandated assessment, who were recruited through professional association newsletters or Toronto-area newspapers. Almost 100 responded to an on-line questionnaire and 40 took part in an hour-long interview; almost 80 per cent of both groups were from the greater Toronto area. Of particular interest were the rationales given by those teachers who said they would not follow the test administration instructions.

#### WHY DOES TEST ADMINISTRATION MATTER?

Research studies have found that how tests are administered affects the results. For example, in a study by Flynn and Anderson (1976), economically disadvantaged grade-7 students in California scored significantly higher on an arithmetic test when the instructions were delivered warmly than when the instructions were delivered perfunctorily. Trentham (1975) found that distractions, including two kittens and a fire alarm, during a test of creativity led the verbal creativity scores of grade-6 students in Kentucky to drop; students' scores in figural creativity, however, were not affected.

The logical conclusion from studies such as these is that standardized administration is important if scores are to be comparable across testing sites. The usual way to encourage standardized test administration is through standardized testing materials, including detailed instructions that must be followed at all sites. These instructions typically specify (a) how students are to be physically arranged in a testing room, (b) when and how the materials are to be distributed and collected, (c) scripts to be read to the students taking the test, (d) what else a person administering the test (the proctor) may and may not say, and (e) the timing of the test sessions.

Following this reasoning, because the instructions are intended to create standardized test administration conditions (and such conditions are essential for the meaningful comparison of test scores across sites), then failure to follow the test administration instructions jeopardizes the comparability of scores. Comparability is not the only concern, however. It is possible to imagine a group of test users agreeing on a new way to administer a test, in which case their results would be comparable with one another. The new administration approach might cause a different problem, however. If the meaning of test scores depends on a test being administered according to the instructions, then the meaning of the scores will be jeopardized.

Most discussions of the importance of following instructions when administering large-scale assessments to students focus on the latter concern. Much research on the effect on score validity of test administration variations relates to accommodations for students with disabilities, such as extended testing time or being read the questions on a mathematics test. Sireci, Scarpati, and Li (2005) describe the rationale for changing test administration procedures for students with disabilities. If some requirements of the usual test administration (e.g., being able to work quickly, having the reading skills to understand mathematics questions) are irrelevant to the meaning of the scores, then changing the administration to remove these requirements for students who have difficulty with them should make the scores for those students more accurate, while not affecting the scores of students who do not have difficulty with the requirements. Based on 28 studies, Sireci et al. concluded that extended time usually leads to better scores for all students and so being able to

work quickly is not easily separable from the knowledge and skills that tests measure. In contrast, having an adult read aloud the questions on mathematics tests provided a greater advantage for students who had been previously identified as having difficulty reading, suggesting that reading skill is separable from mathematics knowledge and skill.

The meaning of the scores is also the focus of discussions about how teachers should prepare students to take tests. For example, Mehrens (1991) writes that "the most general . . . principle is that a teacher should not engage in any type of instruction that attenuates the ability to infer from the test score to the domain of knowledge/skill/ . . . ability of interest" (p. 4). Haladyna, Nolen, and Haas (1991) make a similar point, identifying "nonstandard practices and conditions under which tests are administered" (p. 2), along with some test preparation practices, such as teaching students that if they do not know the right answer to a multiple-choice question they should pick the longest answer, as sources of "test score pollution" because they affect test performance "*without connection to the construct* represented by the test, producing construct-irrelevant test score variance" (p. 4, italics in original).

#### WHEN THE TEST PROCTOR IS THE TEACHER

Several studies have investigated whether teachers, the focus of our study, follow test administration instructions when administering large-scale assessments to their own students. In one of the few observational studies, Wodtke, Harper, Schommer, and Brunelli (1989) coded the actions of kindergarten teachers in 10 classrooms in the United States as they administered standardized tests of school readiness to their students. The teachers varied widely in how closely they followed the instructions, with one teacher providing clues to the correct answer on almost one quarter of her item presentations, and another neglecting to tell the students to close their eyes during a test of recall for which that was a critically important instruction. Wodtke et al., who concluded that four of the 10 teachers made errors in test administration that were large enough to affect the students' test performance or even invalidate one or more of the subtests, acknowledged that the young age of the children might have been a factor in the teachers' test administration behaviors.

Most studies of test administration have simply asked teachers what they or other teachers have done or would do when administering tests. For example, Urdan and Paris (1994) asked 153 kindergarten to grade-8 teachers in the United States about their testing practices and attitudes about testing. Seven per cent said teachers in their school often or always give students help during the administration of a large-scale assessment. In Pedulla, Abrams, Madaus, Russell, Ramos, and Miao's (2003) survey of more than 4,000 elementary and secondary teachers in the United States, teachers reported that the following "unethical test administration practices" had happened in their schools:

1. providing hints (7 to 15%, depending on the test's stakes for schools and students),
2. pointing out mismarked items (8 to 15%),
3. providing more time than allowed (12 to 19%),
4. providing instruction during the test (3 to 9%), and
5. changing students' answers (1 to 2%).

These practices were reported more frequently in states where tests are not used to make decisions about students and schools.

Monsaas and Engelhard (1994) asked 186 teachers, who were taking graduate courses at a Georgia university, whether they engaged in specific practices, including "allowing students extra time to complete the items during the administration of a timed test" and "giving students hints about correct answers while the test was being administered" (p. 472), and whether they believed those practices to be cheating. As they expected, teachers were less likely to report doing things they believed were cheating than things they believed were not cheating.

Kher-Durlabhji and Lacina-Gifford (1992) asked 74 pre-service teachers at a Louisiana university whether they would engage in specific test administration practices and whether they believed those practices were appropriate. Twenty-three per cent believed that rephrasing questions was appropriate, while 34 per cent believed it was inappropriate; 87 per cent believed it was appropriate to check students' completed answer sheets, but none believed it was appropriate to "change answers of low-achieving students" (p. 17). The pre-service teachers were much more likely to predict they would engage in practices that they believed were appropriate than those they believed were not appropriate.

That teachers are not unanimous about which test administration behaviors are ethical and which unethical is also apparent in a study by Green, Johnson, Kim, and Pope (2007). Green et al., who presented testing vignettes to 114 pre-service and 55 in-service teachers at two U.S. universities, asked them whether the teacher in the vignette had acted ethically or unethically. In response to a vignette about a teacher who notices that a student has skipped a line on the answer sheet and reminds him to match the number on the answer sheet to the number of the question, about 70 per cent believed the teacher acted ethically. However, a teacher who notices that a student has answered incorrectly a question he should know and taps her finger by the problem while shaking her head, was thought to be acting ethically by only four per cent. Green et al. conclude that "assessment is currently an educational realm without professional consensus" (p. 999).

#### HOW DO TEACHERS DECIDE WHETHER TO FOLLOW INSTRUCTIONS?

Wellhousen and Martin (1995), in one of the few studies to look more closely at reasons for not following instructions, surveyed 63 pre-service teachers at a U.S. university about whether they would "cheat" when administering a large-scale assessment. Nineteen (30%) of the pre-service teachers reported they would give the students hints or clues during the test. Moreover, four of them viewed this practice as not cheating. When asked if they would do things they believed to be cheating, (a) 6 per cent said they would "if pressured by others" and 6 per cent "to avoid consequences or receive benefits," (b) 20 per cent said they would if they thought "the test or test items were inappropriate" and (c) 36 per cent would "to benefit the children." The reasons for not cheating were similarly varied: (a) 8 per cent would not cheat because "cheating is morally wrong," (b) 5 per cent because of "fear of consequences," and (c) 16 per cent because it would make the results less accurate.

The possibility that the test was inappropriate was echoed by teachers in Smith's (1991) study. Reflecting on her research with elementary teachers in Arizona, Smith observes that some of the teachers believed that the large-scale assessments were so flawed that following the instructions would not yield meaningful scores. She writes:

To chastise teachers for unethical behavior or for “polluting” the inference from the achievement test to the underlying construct of achievement is to miss a critical point: The teachers already view the indicator as polluted. . . . With an interpretive context unavailable to other groups, teachers noted the inadequacies of the mandated achievement test: its poor fit with what they teach, the influence on its scores of pupils’ socioeconomic status and ethnic group, the influence of pupils’ emotional state and intentional effort on test results, its many sources of error, its poor relationship with other indicators of achievement, and its limited scope. (p. 538)

Smith further suggests that teachers’ abilities to closely observe and to empathize with their students may make it difficult for them to ignore the effect of testing on their students.

Wiggins (1994) questions the test administration instructions, but not the appropriateness of the test. He holds that standardizing test administration to ensure “that each student [has] equal opportunity to answer a question correctly . . . hardly justifies the regular practice of forbidding almost all human interaction and the use of contextually appropriate resources, particularly if our aim is to make tests educative and more authentic” (p. 173). Indeed, the test administration instructions, he observes, typically ignore pedagogical considerations.

Green et al. (2007) make similar points in applying to testing the ethical principle of “do no harm”:

There is a potential educational harm done as the result of assessments that fail to accurately measure the knowledge or skills that they claim to measure. There is also the potential emotional harm done to students in the form of anxiety or other stress that high-stakes assessments often bring about. There is also the potential for harm of the teacher-student relationship. Teacher-student trust can be damaged by assessments that the student perceives as unfair or unfounded. (p. 1009)

Green et al. acknowledge that the meaningfulness of the test results is important; however, they also draw attention to a competing concern of many teachers: how students experience testing. In fact, they capture well the tension that exists for teachers between ensuring the meaningfulness of the test results – both as a test proctor by following the admin-

istration instructions, and as a teacher by trying to make sure the students can show what they know – and ensuring that the testing experience is positive for the students and that teacher-student relationships are not jeopardized.

### THIS STUDY

In this study, we asked grade-3 teachers what they would do when faced with test administration dilemmas, and why. We were particularly interested in the teachers' rationales for choosing not to follow the test administration instructions. Grade 3 was chosen, instead of grade 6 or 9, based on anecdotal reports that grade-3 teachers varied more in their test administration practices.

#### *Ontario's Primary Assessment of Reading, Writing, and Mathematics*

Ontario's provincially-mandated *Primary Assessment of Reading, Writing, and Mathematics* (formerly, the *Grade 3 Assessment of Reading, Writing, and Mathematics*) was first administered in 1997. As of 2007, it consists of a mixture of multiple-choice and constructed-response items presented in three test booklets (one containing only mathematics items and two containing a mixture of reading and writing items). Each booklet is administered in two sessions. Each session is typically one hour, but all students are permitted extra time, as long as it is in a continuous sitting.

Detailed instructions that accompany the testing materials include the following statements, which are relevant for this study:

Student assessment material packages must not be opened prior to the administration of the first section of the assessment.

Teachers administering the assessment may receive the student assessment materials no earlier than on the morning of the day the administration begins.

Once the assessment has been opened, no one may use information from the assessment to provide instruction on any concept or item being tested prior to or during the administration of the assessment.

During the assessment, no one may explain, define or provide examples of reading vocabulary or writing or mathematics terminology to students, including those with accommodations.

During the assessment, nothing can be said or done to influence student responses, including, but not limited to, actions such as drawing a student's attention to an unanswered question.



At no point during or after the assessment may anything be said or done to encourage students to alter or revise their responses.

No one may read, review or correct student work during or after the administration. This includes darkening, rewriting, editing, erasing or altering student work in any way. (Education Quality and Accountability Office, 2007, pp. 3-4)

## METHOD

### *Participants*

*Questionnaire.* Between winter and fall of 2007, 117 Ontario elementary teachers completed an on-line questionnaire about what they would do when administering the *Primary Assessment of Reading, Writing, and Mathematics*, and why. The teachers were recruited principally through announcements in their professional association newsletters. All responses were confidential; no honorarium was provided. Of the 117 responses, 98 (11 French, 87 English) that were from teachers who had previously administered the assessment were sufficiently complete for analysis. Of the 98 teachers, 6 were male and 92 female. Twenty of these responses were from teachers outside the greater Toronto area.

*Interviews.* In summer and fall of 2007, the questions from the on-line questionnaire were posed to 40 grade-3 teachers in hour-long, audio-taped interviews (six interviews were conducted in person and the rest by telephone). These teachers were recruited through announcements in their professional association newsletters and through advertisements in community newspapers in the Toronto area. All teachers who volunteered were interviewed, and they received a \$30 bookstore gift certificate. The gift certificates were intended to attract teachers who might not otherwise have been interested in talking about tests. Six of these 40 teachers were from outside the greater Toronto area; four were male.

The 98 responses to the on-line questionnaire and the 40 interviews are the focus of this analysis.

### *Data Collection*

*Developing the vignettes.* The questions used in both the on-line questionnaire and the interviews were developed in the spring of 2005 based on structured interviews with eight grade-3 teachers and three elementary

principals or vice principals; an additional teacher who had formerly taught grade 3 was interviewed in the winter of 2006. Instructors in the University of Toronto's teacher education program nominated these teachers; all teachers and administrators who were nominated were contacted and all agreed to participate. All worked in public or separate (i.e., Catholic) school systems in the greater Toronto area. One teacher was male; the others were female.

We used the responses from these teachers and administrators to create brief vignettes describing dilemmas grade-3 teachers may face in preparing their students for and administering an assessment. This procedure resulted in 10 open-response vignettes, which were pilot tested with several teachers and principals, and then revised. The vignettes were introduced as follows: "In responding to the following situations, imagine you are the teacher of a typical Grade 3 class in an Ontario elementary school." Each vignette ended with the question, "What would you do?" or "What would you say?" For example:

During the test, you walk around the classroom. You notice that one of the students is answering questions with only a word or two. The instructions say students should explain their reasoning. What would you do?<sup>1</sup>

The decision to present the vignettes as hypothetical situations and to ask "What would you do?" instead of "What have you done?" had two motivations. First, asking about what they had done in the past would require that all the teachers had already faced the same dilemmas; we realized that this was unlikely. Second, asking a teacher to report what has happened within her school requires the permission of the school (and, usually, the school board). Obtaining permission from school boards would have made obtaining a wide sample of teachers impractical, if not impossible.

*The questionnaire and interviews.* The resulting questionnaire was posted on a Web site in both English and French. The same questions were used in the hour-long, audiotaped interviews. The interviews were

---

<sup>1</sup> All 10 vignettes are provided in an appendix to this article.

conducted in English by either of the authors, with a second member of the research team also taking notes.

### *Data Analysis*

The audiotapes and notes from the interviews were independently reviewed by a research team member who was not present at the interviews and who subsequently typed summaries of the responses and transcribed relevant quotations. One interview was not audiotaped; that teacher's responses were paraphrased based on detailed notes taken by the interviewer and the observer. The responses to the on-line questionnaires were downloaded and combined with the interview responses. Three team members participated in identifying and coding actions in the combined set of responses, with two members present at each coding session; regular meetings of the entire project team were held to ensure consistency. We took an iterative approach so that once we identified categories and subcategories of actions in one interview, we re-examined other interviews for possible evidence of the same categories and subcategories. For each subcategory within each vignette, we grouped the rationales presented for the actions into themes.

## RESULTS

Comparing what the teachers said they would do in response to the vignettes with the test administration instructions yielded three general categories of actions:

1. noncompliant, in which a teacher chooses to do something that she believes is contrary to the preparation or administration instructions,
2. creatively compliant, in which a teacher chooses to do something that she believes is contrary to the intention of the instructions, but not contrary to a literal interpretation of the instructions, and
3. strictly compliant, in which a teacher follows the instructions.

The results will be organized by *what* the teachers said they would do, while we also explore *why* they said they would take (or not take) specific actions.

*Vignette One: A Mistake in Completing the Answer Sheet*

The teachers were asked:

*As you collect the multiple-choice answer sheets, you notice that one student has coloured two bubbles on several questions. What would you do?*

In response, 64 (46%) said that they would do nothing and so were classified as strictly compliant. The remaining 74 teachers (54%) indicated that they would engage in a combination of noncompliant and creatively compliant actions. The division between noncompliant and creatively compliant is complicated by the need to infer teachers' interpretations of the instructions and the classification is occasionally ambiguous, as illustrated below.

*Noncompliant Actions*

For vignette one, teachers' noncompliant actions fell into two subcategories:

1. pointing out the mistake to the student, and
2. reminding all students to check their answer sheets.

None of the teachers endorsed fixing the student's mistake. Those who mentioned this as a possibility rejected it as unethical.

*Pointing out a mistake to a student.* Teachers who chose to point out a mistake to a student justified this action because poor test-taking skills caused the mistake and the purpose of the test was to measure students' knowledge, not their test-taking skills. Several argued explicitly that, because of this understanding, mistakes caused by test-taking skills would make the test a less accurate measure of a student's knowledge. Answers such as the following illustrate teachers' reasoning:

I would ask them to recheck their answers to make sure they have filled in the bubbles correctly. Correctly filling in a multiple choice answer sheet is not one of the skills that [the test] is assessing. (On-line Survey 11)

Tell them. Bubble sheets are hard to use, even as an adult. I'm not helping them with the answer, and they deserve to show their best. (On-line Survey 71)

*Reminding all students to check their answer sheets.* Some teachers chose not to point out a mistake to a student. As one teacher we interviewed explained it, "Cheating is telling a child 'that's wrong' and singling out

an individual child" (Interview 40). These teachers reasoned, however, that reminding all students to check that they had marked only one answer was ethical. Two teachers who said they would remind the class described how hard it was for them not to point out a mistake to an individual student:

Nothing. It's painful, but you can't do anything. I might do an 'all call' at the end and say, 'Make sure that you have only circled one' . . . (Interview 8)

That's a hard one. I may say: 'Did you remember to choose only one answer?' I would remind the whole class: 'Remember to choose only one answer in multiple-choice questions.' . . . I am not telling them the answers. (On-line Survey 31)

#### *Creatively Compliant Actions*

Creatively compliant actions were of two types:

1. giving instructions before the next session, and
2. reminding a student or all students to read instructions, answer carefully, or do their best work.

*Giving instructions before the next session.* Several teachers pointed out that the test administration instructions did not specify what could be said between testing sessions. Although these teachers did not believe it right to give reminders during a test session, they would instruct students on how to avoid mistakes before the next session. As one teacher, who saw herself as "getting around" the instructions, explained: "During the test itself, teachers are not supposed to talk about it, but there is 10 to 15 minutes before they give the booklets to the students. They can use that time" (Interview 32).

*Reminding all students to read instructions, answer carefully, or do their best work.* When faced with a student who had made a careless mistake on an answer sheet, some teachers said they would give a general reminder about reading the instructions, answering carefully, or doing their best work. They distinguished this type of reminder from a reminder to check the answer sheets for double answers. Many of these teachers believed that the latter would be contrary to the instructions that teachers cannot "encourage students to alter or revise their responses," but that the former would not be contrary to the instructions. For example, one teacher responded: "Nothing. There is nothing I can do

at this point. Just ask if they are 'sure' that they have finished their work and done the best that they can" (On-line Survey 70).

*Strictly Compliant Responses*

To strictly comply with the test administration instructions, a teacher faced with the dilemma in vignette one must do nothing. As expected, many teachers indicated that they could not do anything because that would be contrary to the instructions. One teacher gave this reason for following the instructions:

Although this has not happened, I would like to think I would follow appropriate procedures and not say anything. I am very concerned about how this might affect the overall scores and not following the guidelines might affect the scores. (On-line Survey 98)

However, contrary to our expectations, other teachers gave reasons that were not directly related to the instructions. A few indicated that the typical performance of a student would affect their decision. Interview 18 said, "If [I] knew it was a careless mistake made by a good student then I would say something; otherwise, if this mistake was made by a 'lower end kid,' I would leave it so not to add stress." On-line Survey 16 wrote, "Accept it. Child didn't read carefully. Probably does same thing in class."

The responses of two other teachers suggest that they believed the test was intended to measure the ability to read the instructions, as well as content knowledge:

Just leave it. The instructions are quite clear that you only fill in one bubble. (On-line Survey 65)

No. I know for myself, we did practice multiple-choice and they should know and if they don't know, then they don't know. (Interview 6)

Another teacher pointed out that careless mistakes can be avoided through test preparation:

Nothing. It would be too late. During [the] test prep, I would have taught them how to do [multiple-choice] questions properly, and if they didn't, I would have dealt with it at that time, not during the test. (On-line Survey 46)

### *Vignette Two: Not Writing Enough*

Teachers were also presented the following vignette:

*During the test, you walk around the classroom. You notice that one of the students is answering questions with only a word or two. The instructions say students should explain their reasoning. What would you do?*

The instructions to the provincial tests forbid teachers to say or do anything to “influence student responses” or “encourage students to alter or revise their responses,” including “drawing a student’s attention to an unanswered question” (Education Quality and Accountability Office, 2007, pp. 3-4). In response to this vignette, (a) 46 (33%) of the teachers said they would do nothing and were classified as strictly compliant; (b) 22 (16%) chose only actions that could be classified as creatively compliant; and (c) the remaining 70 teachers (51%) chose noncompliant actions.

#### *Noncompliant Actions*

There were two subcategories of noncompliant actions for this vignette:

1. reminding a student to explain his or her reasoning, and
2. reminding all students to explain their reasoning.

*Reminding a student to explain his or her reasoning.* Many teachers felt that reminding students to explain their answers or to fill the space provided for the answer, while against the instructions, was not ethically wrong. The following responses illustrate their reasoning:

Well, we’re not supposed to say anything to them, but just by telling a student that they need to write more we haven’t changed what they know . . . parents call the principal and complain if the results are poor, so it’s very tempting to intervene in a way that’s morally okay if it will help the results. (Interview 34)

I feel people in the ‘real world’, i.e., outside of the classroom, use tools to help them succeed with a task. When I write a paper, I have someone proof-read it before I hand it in to be judged for a mark. I know my students very well and there are some who need . . . extra reminders to complete something to the best of their abilities. (On-line Survey 52)

*Reminding all students to explain their reasoning.* Some teachers chose to remind the entire class, instead of an individual student, about the

need to explain their answers. One teacher described how she would communicate to her students why explaining answers was important:

I would say, 'Oh, wasn't everybody working really hard?! Let's make sure we put down all the ideas that we have in our heads, because, remember, I am not able to mark the test and the people that mark the test have never met you. So, you need to tell them how intelligent you are . . . ' That is not going to influence the test results. (Interview 32)

This teacher justified her reminder to the class as not unfairly affecting the test results. Another justified her announcement as typical of her usual classroom practice:

I feel that it is important to explain to students as a whole to try their best and that means to add as much important detail as they can or are allowed to do for a question. In a Grade 3 setting I try to maintain some of the regular teacher/student interaction that we are all used to without centering out one student or [giving] away an answer. (On-line Survey 82)

This teacher touches on a concern of many of the teachers: That the pedagogical relationship they had established with students over the course of an academic year included providing encouragement and prompting students to focus and work hard and that they were suddenly being asked to act in ways that were inconsistent with this relationship.

#### *Creatively Compliant Actions*

For vignette two, teachers suggested using three types of creatively compliant actions:

1. reminding a student or all students to read instructions, answer carefully, or do their best work,
2. giving instructions before the next session, and
3. providing nonverbal reminders.

The first two types are the same as for vignette one.

*Reminding a student or all students to read instructions, answer carefully, or do their best work.* Although many teachers saw reminding a student or a class to write more as clearly contravening the instructions, whether it was against the rules to provide more general reminders was less clear.



The following responses emphasize the importance of a student recognizing the need to write more without being told directly:

I would ask the student if this is their best work and something they are proud of and leave the rest up to the individual. (On-line Survey 85)

I would point to the question and tell them to read it again. Because we have discussed this in class all year I would hope that the student understands after that. (On-line Survey 94)

I would ask the child if they are doing their best and encourage the child to continue with a pat on the back. (On-line Survey 26)

*Giving instructions before the next session.* As in their responses to the previous vignette, several teachers pointed out that the test administration instructions did not specify what could be said between testing sessions. Therefore, as illustrated in the following answers, this was a time to remind students about how to take the test:

Before the next section I would remind the class to read instructions carefully and answer questions fully. Even if the question can be answered with only a few words, if they give you 5 lines, write 5 lines. (On-line Survey 78)

After the children have completed and handed in [this part of] their test I ask them how they think they did. I also remind them of the importance of reading the questions and 'filling' the boxes provided! (On-line Survey 90)

*Providing nonverbal reminders.* Although the instructions state that "nothing may be said or *done* to influence student responses" [emphasis added], some teachers saw nonverbal prompts as a way to get around this instruction. The creativity of these teachers' prompts is illustrated in the following responses:

I use a lot of humor with my students, so I would probably give them one of my looks and they would get it. (Interview 11)

Point to the instructions and to the space allotted. The student should be able to understand this prompt. (On-line Survey 87)

I make a reminder board and put it up in front of the class. Then I can point to the reminders but I don't prompt on specific things. (Interview 2)

*Strictly Compliant Actions*

The 33 per cent of the teachers who responded that they would do nothing in response to this vignette offered a myriad of reasons. A few referred to the importance of following the instructions, believing that to intervene would compromise the results: "I would leave them alone. I want my students to earn their level on their own. The children and their parents/guardians need to have a realistic view of what their children have learned" (On-line Survey 89). Other teachers tried to avoid situations such as the one described in the vignette by teaching their students test-taking skills:

As part of any assessment where reasoning was involved, I instructed students how to answer in full sentences and to reread their answers to make sure they would make sense to the reader. I did not instruct students during the test. (On-line Survey 99)

Once the test has started, I do not coach at all. Any student in my class will have heard over 100 times – 'if they give you 10 lines PLEASE do not write two words!!!' Once the test has begun, I am mute! (On-line Survey 53)

Some teachers described vividly how difficult it was for them to follow the instructions:

I would just leave it. It kills me, but I do – that's not what I am there for. (Interview 17)

No, I wouldn't walk around because it's too torturous. I try not to look at what they're writing. (Interview 24)

In our discussion of teachers' choices of actions, we do not suggest that a teacher will always select the same action or that a teacher who selects a noncompliant action in response to one vignette would also do so in response to another – indeed, few of the teachers in our study provided consistently compliant or consistently noncompliant responses across all the vignettes. The following thoughtful response from a teacher who answered the on-line survey illustrates the many factors that would inform her decision in the moment:

My response to that would depend on who the student was. My expectations for my students are that they do the best that they can. It is my expectation throughout the year. If that student is one who really struggles and that is a typical response to short answer questions, I might make a global reminder to everyone in the class to read instructions carefully and make sure you are explaining your answers in detail, to the best of their ability. If the student is one who normally answers questions in complete sentences and with supporting detail, I might ask the student, 'how are you doing?' Students generally reply honestly and you can get a sense as to why their effort is inconsistent with their normal responses. It may be that they had a difficult morning, don't feel well, are tired or feeling stressed, or simply didn't read the question carefully. Their response guides mine. Finally, if all is well with the student and there appears to be no reason for the inconsistency, I might gently remind him/her that short answer questions need to include details that explain their reasoning – or I might congratulate another student, saying, 'I like the way you have a lot of detail in your responses' (thereby supporting one student and encouraging the others). (On-line Survey 84)

This teacher's response provides a window not only into how she decides what to do when administering the test, but also into the pedagogical relationship she has established with her students, her sensitivity to their individual differences, and her understanding of the effect that the testing has on them.

### ***Vignette Three: Prior Knowledge***

A third vignette addressed the use of prior knowledge about the test:

*You work in a school where, when the boxes of tests arrive, the Grade 3 teachers get together and review the test. A question on the test asks students to estimate the volume of a box. You were planning to teach about volume after the test. There's a week before you will be administering the test. What would you do?*

During the study, the test developer increased the clarity of its instructions about access to test materials. This complication was unanticipated. Our intention in writing this vignette was not to find out whether teachers would choose to open the boxes in contravention of the instructions (in fact, we presented the opening of the boxes as something that was not initiated by the teacher), but rather to find out whether they would act on prior knowledge about the test. Opening the boxes is not the only way in which a teacher might gain prior knowledge. For exam-

ple, schools may choose to administer the test on different days within the testing period, so that a teacher might hear about the test from a teacher at another school. Fortunately, in responding to this vignette, most teachers focused on what they would do if they had such knowledge, not on how it was gained. For example, one teacher said:

We are not allowed to look at the test ahead of time. But if I did see the test and notice the volume question, I would review volume. I would feel personally responsible to my students to have taught everything before the test. (Interview 2)

#### *Noncompliant Actions*

In response to the third vignette, 86 (63%) teachers responded that they would teach about volume if they found out that it was the focus of a question on the test. They saw this decision as different from teaching the specific question, which none of the teachers endorsed. The following quotations from three different teachers explain the difference:

I would teach a mini lesson on volume. I know that may sound bad, but if you want the truth, that is what I would do. Grade 3's are little and there is a lot of information on the test. I would like to give them as many opportunities for success as possible. I would not give them exact questions that are on the test (although I worked in a school where such nonsense was done). I think that is wrong, but ensuring they have some knowledge of the concepts is for me an acceptable way to deal with the issue. (On-line Survey 49)

I would take, maybe, one of the days in that week, and try to go over the pages in their textbook relating to volume, along with pages relating to other matters. I'd try not to give undue emphasis to it, and I'd try to present it in a somewhat different format so that I would not be ensuring a 'shoo-in' performance by using [a] similar example. I'd feel responsible, in other words, to cover the material, but I'd want to do that in such a way that I'd not be giving hints and direct examples for them that would be like the test. (On-line Survey 50)

In previous years we would look at the test ahead of time, this year we were only allowed access on the day of testing – so it would be a harder decision. But, I believe I would teach a quick, general lesson about volume. The test is supposed to test the knowledge of the students . . . not trick them because we are a month short on teaching. I know it is an ethical question . . . I would not teach the specif-

ics to answer the question or use anything near the real question – but rather a general approach. The lower and some middle learners will miss a one-lesson overview anyway, but having a question they have no clue how to answer on the test would really stress some children out – especially the top learners. (On-line Survey 53)

All three teachers would teach a quick lesson on volume, but the reasons they gave were different. The first teacher pointed to the students' young age; her concern is to make the testing as positive an experience as possible for these young students by giving them "as many opportunities for success as possible." The second pointed to her responsibility as a teacher to cover the material. The third alluded to the timing of the test – a month before the end of the school year – and to her belief that to encounter a question on unfamiliar material would cause stress for some of her students.

These reasons were echoed in other responses. Several teachers pointed to a sense of duty to their students: They had promised the students that there would not be anything on the test that the students had not been taught. They believed it was very important to honour that commitment to their students and that to do otherwise might jeopardize their relationship with the students. This decision was complicated by the fact that the test covers the entire grade-3 curriculum in mathematics, reading, and writing, although it is administered about a month before the end of the school year. Some teachers see this as placing them in an impossible position. Teachers also made reference to the importance of being fair to their students and, again, the desire to minimize students' stress. For example:

Yes, quite possibly. Otherwise, the kids would say: 'I don't get this! I don't know what this means!' (Interview 30)

Ultimately the problem with the testing . . . is that you're testing on materials that they are expected to have learned, and it is May. Well, theoretically, there's a month, a month and a half left of school . . . if you've not completed that part of the program, is it fair to penalize students for that? . . . Are you teaching to the test? Well, you're going to teach that unit anyway. Why cause them that anxiety? (Interview 5)

I guess I would teach it before the test, as it would not be fair to the students to be presented with unfamiliar material. Students may begin to feel anxiety when faced with a question they don't know. (On-line Survey 39)

I don't think it's fair to test them on stuff they haven't been taught – they would distrust me afterwards. I would not teach them the specific question, just the curriculum expectation. (Interview 16)

A few teachers explicitly referred to higher test scores as their motivation:

I would teach the volume unit because it makes no difference to me when I teach it, so I might as well teach it now to help the students do better on the test. I would like the students to score as well as possible. (Interview 3)

Teach it! The marks come out in the [news]paper. (On-line Survey 77)

### *Creatively Compliant Actions*

There were two subcategories of creatively compliant actions:

1. including the content in a review, and
2. teaching strategies for guessing.

*Including the content in a review.* The teachers who chose to teach the item content understood that they were breaking the rules. For the reasons described above, however, they believed that this action was justified. A smaller group of teachers offered what they saw as a creatively compliant solution: to include volume in the pre-test review. As the following teachers explained, they did not see this as *teaching* the content and so did not see it as violating the instruction that “no one may use information from the assessment to provide instruction on any concept or item being tested.” For example:

I wouldn't teach it, but I would go over a volume question when conducting a broader review on what may be on the test. (Interview 40)

Carry on with doing sample questions with the kids. If I haven't covered volume, I will not teach it, but I might review some of the things they already know about volume and also the strategies that will help them to figure out the answers to a problem like this using what they already know and their reasoning skills. (On-line Survey 70)

*Teaching strategies for guessing.* Another approach a few teachers suggested was to teach the students strategies to guess on questions for which they did not know the answer. Many teachers use this approach. It is creatively compliant only if adopted *after* the teacher learned there would be questions about content that she or he had not taught. For example, one teacher wrote:

I would probably try to find a bit of time to work in some volume lessons, but wouldn't stress about it. I might work with the students on strategies to use when they encounter a question that they are unsure of and would reinforce that they should at least attempt every question. (On-line Survey 43)

#### *Strictly Compliant Actions*

The teachers who said they would not take any of the above actions offered a wide range of reasons. In the following answer, the teacher points to both ethical considerations ("it is wrong") and practical considerations ("this will stress the students out" and "they most likely will not remember"):

This did happen at my school and I disagreed with teaching the students before the test arrived. This will stress the students out and they most likely will not remember what you taught. It is only one question! I think it is wrong and teachers should not be allowed to view the test. (On-line Survey 40)

Another teacher echoed the futility of teaching the content: "I wouldn't teach it because teaching this information last minute won't cause them any deeper understanding or greater recollection of the concepts" (Interview 36). Some teachers explicitly labeled the act of teaching the content as cheating:

No. I think that's cheating. I can't teach to the test. I just can't do it. (Interview 6)  
First, this situation would not happen. I teach the entire content before the test. However, if I was in the situation above, I would not teach volume to the children because it is not fair to the other children in the province who have been prepared appropriately before the test. (On-line Survey 89)

Still others explained that they avoided this situation by making sure they had taught all the curriculum before the test: "That wouldn't happen because we get through the whole math program before the test and then spend June doing all the fun things we missed – cooking, handwriting, etc." (On-line Survey 29).

Although some teachers who endorsed noncompliant actions focused on fairness to the students, other teachers said they would be strictly compliant to be fair to other teachers and schools:

Unlike previous years, one doesn't get to look at the test ahead of time. Hypothetically, it wouldn't be fair to other teachers in the province if that was happening. So, I wouldn't do anything. (Interview 26)

I wouldn't mention this to the students because it is not fair. Some teachers may do this, while others would not. To get valid results, teachers have to follow the instructions. Otherwise, it's really not helping anyone. If it's in the curriculum then the material should have been covered to that point. (Interview 27)

Test materials are not permitted to be reviewed prior to administration so I wouldn't know of the volume question. For these tests to have value we all have to follow the rules. Otherwise our adult politics and press releases hold priority over real benefits to students. (On-line Survey 6)

The complexity of teachers' decision making is illustrated in one teacher's description of the many considerations that would influence her decisions:

Keep everything in perspective and know your kids. The question on the test about the volume of the box is only one question out of many. If students haven't been taught that formally yet this year, one question will not significantly change their entire test results. Further, students can also draw upon previous knowledge, presented or learned in other grades or environments. However, that also doesn't mean that I definitely would not teach volume to my students. It would depend on whether or not they are at a stage where learning that information is something that takes priority over other planned activities and whether or not they are at a developmental level where concise presentation of the concepts would promote understanding or more confusion. (On-line Survey 84)



## DISCUSSION

That many teachers predict they would not strictly follow test administration instructions is hardly surprising. Numerous studies (e.g., Green et al., 2007; Kher-Durlabhji & Lacina-Gifford, 1992; Monsaas & Engelhard, 1994; Pedulla et al., 2003; Urdan & Paris, 1994; Wodtke et al., 1989) have asked teachers *what* they have done or would do and found a wide variation of responses. With the exception of Wellhousen and Martin (1995), however, few researchers have investigated *why*.

This study is limited by the necessity to ask teachers to volunteer to participate; for this reason it is impossible to generalize about the prevalence of the represented views in the general population. Nonetheless, it provides the most detailed descriptions to date of how teachers decide what to do in testing situations. Furthermore, the range of responses is persuasive in indicating the diverse actions teachers predict they would take.

To understand the decisions of the teachers in this study, it is useful to consider two focuses in combination with the dual roles of teacher and test proctor: the test results and the testing experience.

In their role as test proctor, some teachers were motivated to comply with the test administration instructions by concern for the comparability of the test results – if the tests were administered differently in different classrooms, they worried, the results might not be comparable. Some teachers also believed that following the instructions was important for the accuracy of the results – that is, the tests would not mean what they were intended to mean if the instructions were not strictly followed.

In their role as classroom teacher, however, the respondents' focus on the test results was related to a worry that strictly following the test administration instructions might make the test results *less* accurate. This depended, of course, on what the teachers believed the test was intended to measure. With respect to the first two vignettes, some teachers believed that the test should measure what students know, while others believed it should measure what students know *plus* their skill in taking tests. For the third vignette, teachers differed in whether they believed the test was meant to measure students' ability or their knowledge.

Many teachers who chose not to follow the instructions because of concern about the test results did so to help students with weak test-

taking skills or who were nervous or distracted during the test. Consistent with Smith's (1991) analysis, these teachers believed they were correcting flaws in the test itself, such as confusing wording, or in the test instructions, such as inappropriate restrictions on encouraging the students. These teachers believed they were helping their students to "show what they know." Because they were not helping the students to "show what they *don't* know," they did not consider it cheating. They believed they were increasing the meaningfulness of the test scores.

The results of the tests, however, were only one consideration for many of the teachers in this study. They were also worried about the testing experience for their students. In the test proctor role, they had little opportunity to consider how the students experienced taking the tests; indeed, the test proctor's responsibility extends only to controlling what is said and done in the test setting, not to the test takers' subjective experience. For most teachers, however, the students' experience was very important. As Green et al. (2007) write, testing has the potential to cause some students extreme anxiety and even to harm student-teacher relationships. Many teachers in this study described wanting to minimize the stress of testing for their students. Whether they were willing to contravene the test administration instructions to do this depended in part on their beliefs about the meaningfulness of the test results. As Smith (1991) suggests, teachers who believe the test was irrelevant may shift their focus from worrying about the accuracy of the test results to minimizing the stress of the testing experience. Wellhousen and Martin (1995) also found that teachers' beliefs about the appropriateness of a test affected their decisions.

Although defining decisions about test administration as compliant or not is central to this study, equally important is whether teachers viewed their decisions as ethical. Test developers have typically tried to encourage compliance by exhorting teachers to act ethically. If, however, teachers do not believe that their decisions, whether compliant or not, are unethical, then such exhortations are unlikely to be effective. In general, the teachers in this study presented their decisions either as in the best interest of the students (and, therefore, justified by their role as a teacher) or else as in strict compliance with instructions (and, so, justified

by their role as a test proctor). That many decisions were consistent with only one of the roles was not necessarily seen as a threat to ethicality.

Considering how the roles of test proctor and teacher interact with the focuses on test results and testing experience makes it clear why the testing situations examined in this study present dilemmas. Concern about the comparability or accuracy of test results may suggest different decisions than would a focus on the students' testing experience. This conclusion is complicated by the two roles, which may assign different priorities to and have different views of test results and testing experience. That different teachers reach different decisions – and that individual teachers may change their decisions based on the specific tests and students involved – reflects the complexity of this interaction among roles and focuses.

#### SUMMARY

When faced with a student who has made a careless mistake in filling in an answer sheet (vignette one) or is writing very short responses to questions requiring explanation or further description (vignette two), teachers administering Ontario's provincially mandated assessment to their grade-3 students are required to say and do nothing. However, many teachers in this study predicted that, if faced with these situations, they would not strictly comply with the test administration instructions. Instead, they would either act noncompliantly – that is, in a way that clearly violates the instructions – or creatively comply by doing something they believed circumvented the rules. Even more strikingly, if they found out before the test that it would include something they had not yet taught their students, more than half the teachers would quickly teach the missing content. The reasons teachers gave for their decisions reflect an interaction between the sometimes competing roles of test proctor and teacher, with two focuses: on the test results and on how the students experience the testing. In particular, teachers who believed the administration instructions were inappropriate for grade-3 students did not view failing to follow the instructions as unethical.

*Implications*

The findings suggest that test developers need to consider the role of the teacher when they design the role of the test proctor. Teachers have spent the academic year getting to know their students' individual needs and characteristics and developing a pedagogical relationship based on trust. They are committed to helping students do their best work and have developed a wide range of strategies to encourage, prompt, or cue students in their everyday classwork and classroom tests. They want their students to have a positive testing experience. If test developers want teachers to follow test administration instructions, they need to consider the potential conflicts between this complex role and the test proctor role when they design the instructions.

**CONCLUSION**

This study has several implications for organizations that create and oversee large-scale assessments. If teachers are choosing not to follow the test administration instructions, then the comparability and accuracy of the test results may well be compromised. However, trying to standardize test preparation and administration by insisting that teachers follow instructions is likely to be only partially successful if teachers believe that not following the instructions is actually more consistent with their professional obligations as teachers. Because many teachers in our study believed the current instructions were unreasonable – and, indeed, showed the provincial testing agency to be out of touch with teachers and students – they did not perceive violating the instructions as unethical.

It may be that the requirement that teachers follow the test instructions as currently written when administering the test to their own students is unrealistic. One possible solution is to develop test administration instructions that are consistent with teachers' primary roles. For example, teachers might be permitted to have more interactions with students, such as encouraging students to do their best and to stay focused on the task. Large posters with test-taking tips might be distributed well before the testing and used for review in the months leading up to, as well as immediately before a test. Such changes would not eliminate variation among teachers in test administration, but might reduce the varia-

tion. However, as long as the instructions continue to present professional dilemmas for teachers, different teachers will choose different practices.

#### ACKNOWLEDGEMENTS

Thank you to the teachers who participated in this research and to Madison Aitken, Susan Elgie, Lina Fung, Monique Herbert, Vesna Milenkovic, Anita Ram, and Yunmei Xu for invaluable assistance with the data collection and analysis. We gratefully acknowledge the support of the Social Sciences and Humanities Research Council of Canada for this research.

#### REFERENCES

- Cizek, G. J. (1999). *Cheating on tests: How to do it, detect it, and prevent it*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Cizek, G. J. (2003). Educational testing integrity: Why educators and students cheat and how to prevent it. In J. Wall & G. R. Walz (Eds.), *Measuring up: Assessment issues for teachers, counselors, and administrators* (pp. 363-386). Greensboro, NC: CAPS Press.
- Cohen, S. A., & Hyman, J. S. (1991). Can fantasies become facts? *Educational Measurement: Issues and Practice*, 10(1), 20-23.
- Crocker, L. (2006). Preparing examinees for test taking: Guidelines for test developers and test users. In S. M. Downing & T. Haladyna (Eds.), *Handbook of test development* (pp. 115-128). Mahwah, NJ: Lawrence Erlbaum Associates.
- Education Quality and Accountability Office (EQAO). (2007). *Administering the Assessments of Reading, Writing and Mathematics. Primary Division (Grades 1-3) and Junior Division (Grades 4-6). Spring 2007*. Toronto: Author.
- Flynn, J. T., & Anderson, B. E. (1976). The effect of test administration procedures on achievement test performance. *Journal of Negro Education*, 45(1), 37-45.
- Green, S. K., Johnson, R. L., Kim, D.-H., & Pope, N. S. (2007). Ethics in classroom assessment practices: Issues and attitudes. *Teaching and Teacher Education*, 23(7), 999-1011.

- Haladyna, T. M., Nolen, S. B., & Hass, N. S. (1991). Raising standardized achievement test scores and the origins of test score pollution. *Educational Researcher*, 20(5), 2-7.
- Kher-Durlabhji, N., & Lacina-Gifford, L. J. (1992, November). *Quest for test success: Preservice teachers' views of "high stakes" tests*. Paper presented at the annual meeting of the Mid-South Educational Research Association, Knoxville, TN. Retrieved July 22, 2009, from the Educational Resources Information Center (ERIC) Web site (ERIC Document Reproduction Service No. ED353338): [http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content\\_storage\\_01/0000019b/80/13/3e/62.pdf](http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/13/3e/62.pdf)
- Mehrens, W. A. (1991, April). *Defensible/indefensible instructional preparation for high stakes achievement tests: An exploratory dialogue*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL. Retrieved July 22, 2009, from the Educational Resources Information Center (ERIC) Web site (ERIC Document Reproduction Service No. ED334202): [http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content\\_storage\\_01/0000019b/80/23/11/ad.pdf](http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/23/11/ad.pdf)
- Mehrens, W. A., & Kaminski, J. (1989). Methods for improving standardized test scores: Fruitful, fruitless, or fraudulent. *Educational Measurement: Issues and Practice*, 8(1), 14-22.
- Mehrens, W. A., Popham, W. J., & Ryan, J. M. (1998). How to prepare students for performance assessments. *Educational Measurement: Issues and Practice*, 17(1), 18-22.
- Monsaas, J., & Engelhard, G., Jr. (1994). Teachers' attitudes toward testing practices. *Journal of Psychology*, 128(4), 469-477.
- Pedulla, J. J., Abrams, L. M., Madaus, G. F., Russell, M. K., Ramos, M. A., & Miao, J. (2003). *Perceived effects of state-mandated testing programs on teaching: Findings from a national survey of teachers*. Chestnut Hill, MA: National Board on Educational Testing and Public Policy. Retrieved July 22, 2009, from the Educational Resources Information Center (ERIC) Web site (ERIC Document Reproduction Service No. ED481836):

[http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content\\_storage\\_01/0000019b/80/1b/80/96.pdf](http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/1b/80/96.pdf)

- Popham, W. J. (1991). Appropriateness of teachers' test-preparation practices. *Educational Measurement: Issues and Practice*, 10(4), 12-15.
- Sireci, S. G., Scarpati, S. E., & Li, S. (2005). Test accommodations for students with disabilities: An analysis of the interaction hypothesis. *Review of Educational Research*, 75(4), 457-490.
- Smith, J. L. (1991). Meanings of test preparation. *American Educational Research Journal*, 28(3), 521-542.
- Trentham, L. L. (1975). The effect of distractions on sixth-grade students in a testing situation. *Journal of Educational Measurement*, 12(1), 13-17.
- Urdu, T. C., & Paris, S. G. (1994). Teachers' perceptions of standardized achievement tests. *Educational Policy*, 8(2), 137-156.
- Wellhousen, K., & Martin, N. K. (1995). Preservice teachers and standardized test administration: Their behavioral predictions regarding cheating. *Research in the Schools*, 2(2), 47-49.
- Wiggins, G. (1994). The immorality of test security. *Educational Policy*, 8(2), 157-182.
- Wodtke, K. H., Harper, F., Schommer, M., & Brunelli, P. (1989). How standardized is school testing? An exploratory observational study of standardized group testing in kindergarten. *Educational Evaluation and Policy Analysis*, 11(3), 223-235.

*Ruth Childs* is a professor in the Department of Human Development and Applied Psychology, Ontario Institute for Studies in Education, University of Toronto.

*Linda Umezawa* is a graduate student in the Developmental Psychology and Education Program in the Department of Human Development and Applied Psychology, Ontario Institute for Studies in Education, University of Toronto.

Contact information:

Ruth Childs   rchilds@oise.utoronto.ca

Ontario Institute for Studies in Education/ University of Toronto

252 Bloor Street West, 9th Floor

Toronto, ON M5S 1V6   (416) 978-1079

## Appendix

*In responding to the following situations, imagine you are the teacher of a typical Grade 3 class in an Ontario elementary school. Your students will be taking Ontario's Grade 3 Assessment of Reading, Writing and Mathematics in May.*

1. A week after school begins in September, a parent phones you to talk about the test. She is very worried that writing the Grade 3 Assessment of Reading, Writing and Mathematics may be stressful for her child. What would you tell her?
2. At a workshop about the test, the presenter suggests that you teach your students strategies for answering multiple-choice questions (for example, the students should eliminate the choices that they think are wrong, before picking an answer). Would you follow this suggestion? Why or why not?
3. A new Grade 3 teacher comes to you for advice about how to prepare his class for the test. Which of the following activities or approaches do you recommend he use? (Note: The response options were Yes, definitely; Maybe; No, definitely not; and Unsure.)
  - Teaching strategies for answering multiple-choice questions (e.g., narrowing the choices, filling in the bubbles on the scan sheets completely, guessing)
  - Teaching strategies for answering open response questions (e.g., planning their responses, explaining their responses, writing responses in the space provided)
  - Teaching how to understand the test instructions
  - Creating classroom tests that use the question formats that will be used on the test
  - Creating classroom tests that use instructions similar to those that will be used on the test
  - Helping students get used to working independently
  - Talking about the importance of the test
  - Talking about the unimportance of the test
  - Teaching students how to handle feelings of anxiety about the test



- Having students practice answering sample questions during class, but under relaxed conditions
  - Having students work in small groups on sample questions
  - Discussing examples of good responses to sample questions
  - Administering sample questions under conditions like the real test
  - Sending a letter to parents about how they can help their children prepare for the test
4. In the staff lounge, two teachers are arguing about the test. One believes the test administration itself takes too much time and refuses to give another minute to the test. She doesn't talk about the test with her students or give them examples. The other teacher argues that talking with the students about the test and helping them become familiar with the format decreases the students' anxiety. They ask you what you think. What would you say?
  5. You work in a school where, when the boxes of tests arrive, the Grade 3 teachers get together and review the test. A question on the test asks students to estimate the volume of a box. You were planning to teach about volume after the test. There's a week before you will be administering the test. What would you do?
  6. During the test, you walk around the classroom. You notice that one of the students is answering questions with only a word or two. The instructions say students should explain their reasoning. What would you do?
  7. During the test, you notice that one of your strongest students is spending too much time on the first few questions. You are worried that she may run out of time later in the test. What would you do?
  8. During the test, you notice that one of your students has made a careless error in the first step of a multi-step math problem. What would you do?
  9. As you collect the multiple-choice answer sheets, you notice that one student has coloured two bubbles on several questions. What would you do?
  10. Your vice principal compares the EQAO results from last year with your report card marks for those same students last year. She tells you that your report card marks do not agree with the EQAO results

and suggests that you need to be more careful in your marking.  
What would you say?